Filamentous fungal mutants with improved homologous recombination efficiency

Field of the invention

The invention relates to the field of molecular biology. It particularly relates to methods to improve the efficiency of directed integration of nucleic acids into the genome of a filamentous fungus and uses thereof.

Background of the invention

Eukaryotic cells are preferred organisms for (recombinant) production of polypeptides and secondary metabolites. When constructing, for example, a protein production strain, the site of integration of the gene of interest coding for the protein to be produced is crucial for the regulation of transcription and/or expression of the integrated gene of interest. Since in most eukaryotic organisms integration of DNA into the genome occurs with high frequency at random, the construction of a protein production strain by recombinant DNA technology often leads to the unwanted random integration of the expression cassette comprising the gene encoding the protein to be produced. This uncontrolled "at random multiple integration" of an expression cassette is a potentially dangerous process, which can lead to unwanted modification of the genome of the host. It is therefore highly desirable to be able to construct a protein production strain by ensuring the correct targeting of the expression cassette with high efficiency.
Furthermore, now that the sequence of complete genomes of an increasing amount of organisms is becoming available, this opens the opportunity to construct genome spanning overexpression and deletion libraries. An important requirement for the efficient construction of such libraries is that the organism in question can be efficiently transformed and that the required homology needed to direct targeted integration of a nucleic acid into the genome is relatively short.

Eukaryotic cells have at least two separate pathways (one via homologous and one via non-homologous recombination) through which nucleic acids (in particular of course DNA) can be integrated into the host genome. The yeast *Saccharomyces cerevisiae* is an organism with a preference for homologous recombination (HR). The ratio of non-

homologous to homologous recombination (NHR/HR) of this organism may vary from about 0.07 to 0.007.

WO 02/052026 discloses mutants of *Saccharomyces cerevisiae* having an improved targeting efficiency of DNA sequences into its genome. Such mutant strains are deficient in a gene involved in NHR (KU70).

Contrary to *Saccharomyces cerevisiae*, most higher eukaryotes such as filamentous fungal cells up to mammalian cell have a preference for NHR. Among filamentous fungi, the NHR/HR ratio is ranged between 1 and more than 100. In such organisms, targeted integration frequency is rather low. To improve this frequency, the length of homologous regions flanking a polynucleotide sequence to be integrated into the genome of such organisms has to be relatively long for example at least 2000bp for disrupting a single gene and at least 500bp for screening putative transformants. The necessity of such flanking regions represents a heavy burden when cloning the DNA construct comprising said polynucleotide and when transforming the organism with it. Moreover, neighbouring genes which lie within those flanking regions can easily be disturbed during the recombination processes following transformation, thereby causing unwanted and unexpected side-effects.

Mammalian cells deficient in KU70 have already been isolated (Pierce et al, Genes and Development, (2001), 15: 3237-3242). These mutants have a six-fold higher homology-directed repair frequency, but no increase in the efficiency of homology-directed targeted integration. This suggests that results obtained in organisms with a preference for HR (*Saccharomyces cerevisiae*) cannot be extrapolated to organisms with a preference for NHR.

Surprisingly, we found that steering the integration pathways of nucleic acids towards HR in filamentous fungi resulted in an improved efficiency for targeted integration of nucleic acids into the genome of filamentous fungi.

Brief description of the drawings

Figure 1 depicts the replacement vector pDEL-HDFA used to inactive the *hdfA* gene in *Aspergillus niger (A. niger)*. The replacement vector comprises the *hdfA* flanking regions,

the *amdS* marker and *E. coli* DNA. The *E. coli* DNA was removed by digestion with restriction enzymes AscI and NotI, prior to transformation of the *A. niger* strains.

Figure 2 depicts the replacement vector pDEL-HDFB used to inactive the *hdfB* gene in *A. niger*. The replacement vector comprises the *hdfB* flanking regions, the *amdS* marker and *E. coli* DNA. The *E. coli* DNA was removed by digestion with restriction enzymes AscI and NotI, prior to transformation of the *A. niger* strains.

Figure 3 depicts the strategy used to delete the *hdfA* gene of *A. niger*. The DNA construct used comprises the *amdS* selection marker flanked by homologous regions (5' and 3') of the *hdfA* gene (1). This construct integrates through double homologous recombination (X) at the genomic *hdfA* locus (2) and replaces the genomic *hdfA* gene copy (3). Subsequently, recombination over the direct repeats (U) removes the *amdS* marker, resulting in precise excision of the *hdfA* gene (4).

Figure 4 depicts the strategy used to delete the *hdfB* gene of *A. niger*. The DNA construct comprises the *amdS* selection marker flanked by homologous regions (5' and 3') of the *hdfB* gene (1). This construct integrates through double homologous recombination (X) at the genomic *hdfB* locus (2) and replaces the genomic *hdfB* gene copy (3). Subsequently, recombination over the direct repeats (U) removes the *amdS* marker, resulting in precise excision of the *hdfB* gene (4).

Figure 5 depicts the schematic strategy used to integrate a DNA construct into the genome of *A. niger* through single homologous recombination. The expression vector comprises the selectable *amdS* marker and a gene of interest flanked by homologous regions of the *glaA* locus (3' *glaA* and 3" *glaA* respectively) to direct integration at the genomic *glaA* locus.

Description of the invention

All patents and publications, including all sequences and methods disclosed within such patents and publications, referred to herein are expressly incorporated by reference. These patents and publications include: EP 357 127 B, EP 635 574 B, WO 97/06261, WO 98/46772.

Method for increasing the efficiency of targeted integration of a polynucleotide into the genome of a filamentous fungal cell

The present invention provides a method for increasing the efficiency of targeted integration of a polynucleotide to a pre-determined site into the genome of a filamentous fungal cell, with a preference for NHR, wherein said polynucleotide has a region of homology with said pre-determined site comprising steering an integration pathway towards HR. The present invention arrives at such steering either by elevating the efficiency of the HR pathway, and/or by lowering (meaning reducing) the efficiency of the NHR pathway and/or by decreasing the NHR/HR ratio.

In the context of the invention, the HR pathway is defined as all genes and elements being involved in the control of the targeted integration of polynucleotides into the genome of a host, said polynucleotides having a certain homology with a certain pre-determined site of the genome of a host wherein the integration is targeted. The NHR pathway is defined as all genes and elements being involved in the control of the integration of polynucleotides into the genome of a host, irrespective of the degree of homology of the said polynucleotides with the genome sequence of the host.

According to a preferred embodiment, the steering comprises providing a mutant of a parent filamentous fungal cell, wherein the NHR/HR ratio is decreased in the mutant of at least 5% as compared to said ratio in said parent organism as measured by the following assay. More preferably, the NHR/HR ratio is decreased in the mutant of at least 10%, even more preferably at least 50% and most preferably at least 100% as compared to said ratio in said parent organism.

According to another preferred embodiment, the filamentous fungal cell of the invention has a ratio NHR/HR, which is at least 200, at least 50, at least 10 as measured by the following assay. Preferably the ratio of the filamentous fungal cell is at least 1, more preferably at least 0.5, even more preferably at least 0.1, even more preferably at least 0.05, even more preferably at least 0.01 even more preferably at least 0.005 even more preferably at least 0.001 even more preferably at least 0.0005 even more preferably at least 0.0001 and most preferably at least 0.00001.

According to a more preferred embodiment, the filamentous fungal cell of the invention has a ratio NHR/HR, which is less than 200, even more preferably less than 50, less than 10 as measured by the following assay. Even more preferably the ratio of the

filamentous fungal cell is less than 1, even more preferably less than 0.5, even more preferably less than 0.1, even more preferably less than 0.05, even more preferably less than 0.0 1 even more preferably less than 0.005 even more preferably less than 0.001 even more preferably less than 0.0005 even more preferably less than 0.0001 and most preferably less than 0.00001.

The ratio of NHR/HR is preferably measured by the assay as described in WO 02/052026 (table 2, p23). According to a preferred embodiment, the parent organism is one of the filamentous fungus cells as defined under the section host cell. According to another preferred embodiment, the filamentous fungus cell of the invention originates from a species as defined under the section host cell.

Alternatively and according to a less preferred embodiment, the NHR/HR ratio in a filamentous fungus is monitored using techniques known to the skilled person such as transcriptional profiling and/or northern blotting and/or western blotting of at least one of the following components involved in such pathways: KU70, KU80, MRE11, RAD50, RAD51, RAD52, XRS2, SIR4, LIG4.

In the context of this invention, "a region of homology" means "at least one" region of homology. A pre-determined site is herein defined as a site within the genetic material contained by a host cell to which a polynucleotide with homology to this same site is integrated with a method according to the invention.

In a preferred embodiment, the invention provides a method for increasing the efficiency of targeted integration of a polynucleotide to a pre-determined site into the genome of a filamentous fungal cell with a preference for NHR, wherein said polynucleotide has a region of homology with said predetermined site comprising steering an integration pathway towards HR by providing a filamentous fungus, wherein the efficiency of the NHR pathway has been lowered and/or the NHR/HR ratio has been decreased compared to the efficiency of the NHR pathway and/or the NHR/HR ratio of the filamentous fungus it originates from under the same conditions. According to a preferred embodiment, the parent organism is one of the filamentous fungus as defined under the section host cell.

The efficiency of the NHR pathway is preferably measured in the assay as described in WO02/052026 (table 2, p23).

Alternatively and according to a less preferred embodiment, the efficiency of the NHR pathway in a filamentous fungus is monitored using techniques known to the skilled person such as transcriptional profiling and/or northern blotting and/or western blotting of components involved in such pathway. More preferably, the expression level of at least one of the following components is monitored: KU70, KU80, MRE11, RAD50, RAD51, RAD52, XRS2, SIR4, LIG4. Even more preferably, the expression level of homologous components of the KU complex is monitored. Most preferably, the expression level of homologous KU70 and/or KU80 is monitored.

A lowered NHR efficiency means at least lower than in the parental cell the obtained cell originates from. Preferably, lowered means twice lower, more preferably ten times lower, even more preferably 100 times lower, most preferably more than 1000 times lower and even most preferably not detectable using northern or western blotting, array techniques or a phenotypic screen.

A typical phenotypic screen that could be used comprises the following steps: transforming the putative NHR mutants with an expression cassette comprising a selection marker gene flanked by homologous sequences of a predetermined genomic site. The selection marker gene used in this phenotypic screen can be selected from a number of marker genes that are useful for transformation of filamentous fungi. By way of example these markers include but are not limited to dominant and bi–directional selection marker gene such as an acetamidase (*amdS*) gene (EP 635 574 B or WO 97/06261), auxotrophic marker genes such as *argB*, *trpC* or *pyrG* and antibiotic resistance genes providing resistance against e.g. phleomycin (the product encoded by the *ble* gene confers resistence to phleomycine), hygromycin B or *G418*. A preferred selection marker gene is the *ble* gene encoding a protein conferring resistence to phleomycin. The putative NHR mutants already contain at this predetermined genomic site a birectional selection marker gene such as an *amdS* gene, nitrate reductase gene (*niaD*), sulphate permease (*Sut B*) gene or *PyrG* gene. The *niaD* gene has already been described elsewhere (Gouka RJ, van Hartingsveldt W, Bovenberg RA, van den Hondel CA, van Gorcom RF. Cloning of the nitrate-nitrite reductase gene cluster of Penicillium chrysogenum and use of the *niaD* gene as a homologous selection marker. J Biotechnol. 1991 Sep;20(2):189-99). The *niaD* gene enables direct selection of transformants on

plates containing chlorate, as cells become resistant to chlorate. The *sutB* gene has already been described elsewhere (van de Kamp M, Pizzinini E, Vos A, van der Lende TR, Schuurs TA, Newbert RW, Turner G, Konings WN, Driessen AJ. Sulfate transport in Penicillium chrysogenum: cloning and characterization of the *sutA* and *sutB* genes. J. Bacteriol. 1999 Dec;181(23):7228-34). A preferred selection marker gene is the *A.nidulans amdS* coding sequence fused to the *A.nidulans gpdA* promoter (EP635 574 B). *AmdS* genes from other filamentous fungi may also be used (WO 97/06261). In the preferred form of the phenotypic screen, the *amdS* gene is present at the predetermined genomic site and the *ble* gene is used as the gene to be targeted to the predetermined site. In non-HR-improved mutants the *ble*-cassette will integrate randomly in the genome, enabling many transformants to grow on a double selective medium with both acetamide and phleomycin; and relatively few transformants to grow on fluoracetamide-phleomycin plates. In mutants with improved HR there will be a limited number of transformants on the acetamide–phleomcycin double selective plates as the *amdS*-cassette is efficiently exchanged with the *ble*-cassette. In this case more mutants will appear on fluoracetamide-phleomycin double selective plates.

According to another preferred embodiment, the filamentous fungus having a lowered NHR efficiency and/or a decreased NHR/HR ratio is a filamentous fungus wherein a component involved in NHR has been inhibited. In this context, "a" means "at least one": at least one component involved in NHR has been inhibited in a given filamentous fungus. Inhibition can be achieved by down regulating the expression level of a gene involved in NHR or inactivating a gene encoding a component involved in NHR and/or by down regulating the expression level of a component involved in NHR, and/or (temporarily) decreasing the (protein) activity of a component involved in NHR and a combination of these possibilities.

Preferably, the filamentous fungus obtained has the expression of a gene involved in NHR down regulated by comparison to the expression of said gene in the parent filamentous fungal cell it originates from under the same conditions. According to a preferred embodiment, the parent filamentous fungus is one of the filamentous fungus as defined under the section host cell.

The expression level of a gene, or a DNA sequence is down regulated when the expression level of this specific gene or DNA sequence in the obtained filamentous fungus is lower than the expression level of the same gene or DNA sequence in the

parental filamentous fungus it originates from, preferably three times lower, more preferably four times lower, most preferably more than four times lower and even most preferably not detectable using northern, or western blotting or 'omics' techniques like transcriptomics and proteomics.

The down and/or up regulation of the expression level of a DNA sequence can be monitored by quantifying the amount of corresponding mRNA present in a cell by northern blotting (in Molecular Cloning: A Laboratory Manual, Sambrook *et al.*, New York: Cold Spring Harbour Press,1989) for example and/or by quantifying the amount of corresponding protein present in a cell by western blotting for example. The difference in mRNA amount may also be quantified by DNA array analysis (Eisen, M.B. and Brown, P.O. DNA arrays for analysis of gene expression. Methods Enzymol. 1999:303:179-205).

The down regulation of the expression level of at least one gene or DNA sequence may be obtained by genetic manipulation by one of the following techniques or by a combination thereof:

a. using recombinant genetic manipulation techniques,
b. submitting the filamentous fungus to mutagenesis.

Alternatively or in combination with above-mentioned techniques and according to another preferred embodiment, the down regulation of the expression level of at least one gene or DNA sequence may be obtained by submitting the filamentous fungus to a inhibiting compound/composition.

The filamentous fungus obtained may be subsequently selected by monitoring the expression level of said gene or DNA sequence. Optionally, the filamentous fungus is subsequently selected by measuring its efficiency of the NHR and/or of the HR pathways and/or its NHR/HR ratio. In the context of the invention, the efficiency of the HR pathway of a filamentous fungus may be measured by the efficiency of the targeted integration of a given poynucleotide sequence into a pre-determined site in the genome of the filamentous fungus using given homology region(s). In the context of the invention, the efficiency of the NHR pathway of a filamentous fungus may be measured by the efficiency of the non targeted integration of a given polynucleotide sequence in the genome of the filamentous fungus irrespective of any homology region(s).

More preferably, the down regulation of the expression of at least one DNA sequence is made with recombinant genetic manipulation techniques such as defined in step a. to obtain a recombinant filamentous fungus. Most preferably step a. comprises deleting the DNA sequence, even most preferably the deleted DNA sequence is replaced by a non-functional variant thereof, and even most preferably the deletion and replacement are made by gene replacement preferably as described in EP 357127 B.

In cases of deletion or replacement of at least one DNA sequence of the chosen filamentous fungus, an appropriate DNA sequence has to be introduced at the target locus. The target locus is in this case the DNA sequence involved in NHR pathway to be deleted or replaced. The appropriate DNA sequence is preferably present on a cloning vector. Suitable cloning vector are the ones that are able to integrate at the pre-determined target locus in the chromosomes of the filamentous fungal host cell used. Preferred integrative cloning vector comprises a DNA fragment, which is homologous to the DNA sequence to be deleted or replaced for targeting the integration of the cloning vector to this pre-determined locus. In order to promote targeted integration, the cloning vector is preferably linearized prior to transformation of the host cell. Preferably, linearization is performed such that at least one but preferably either end of the cloning vector is flanked by sequences homologous to the DNA sequence to be deleted or replaced.

The length of the homologous sequences flanking the DNA sequence to be deleted or replaced is preferably less than 2 kb, even preferably less than 1 kb, even more preferably less than 0.5kb, even more preferably less than 0.2kb, even more preferably less than 0.1kb, even more preferably less than 50bp and most preferably less than 30bp.

The selection marker gene in the cloning vector can be selected from a number of marker genes that are useful for transformation of filamentous fungi. By way of example these markers include but are not limited to dominant and bi-directional selection marker gene such as an acetamidase (*amdS*) gene (EP 635 574 B or WO 97/06261), auxotrophic marker genes such as *argB*, *trpC*, or *pyrG* and antibiotic resistance genes providing resistance against e.g. phleomycin, hygromycin B or G418. A preferred selection marker gene is the *A.nidulans amdS* coding sequence fused to the *A.nidulans gpdA* promoter (EP635 574 B). *AmdS* genes from other filamentous fungus may also be used (WO 97/06261). The *amdS* selection marker gene has the advantage it can be used several times in the same strain to replace and/or delete distinct DNA sequences.

By means of counterselection on fluoracetamide media as described in EP 635 574 B, the resulting strain is marker free and can be used for further gene modifications.

A preferred strategy for down regulating the expression of a given DNA sequence comprises the deletion of the wild type DNA sequence and/or replacement by a modified DNA sequence, whose expression product is not functional. The deletion and the replacement are preferably performed by the gene replacement technique described in EP 0 357 127 B1.The specific deletion of a gene is preferably performed using the *amdS* gene as selection marker gene as described in EP 635 574 B.

Alternatively or in combination with other mentioned techniques, a technique based on *in vivo* recombination of cosmids in *E. coli* can be used, as described in: A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans* (2000) Chaveroche, M-K., Ghico, J-M. and d'Enfert C; Nucleic acids Research, vol 28, no 22. This technique is applicable to other filamentous fungi like for example *A. niger*.

Down regulating the expression of a DNA sequence may also be achieved by using anti sense nucleic acids, or by UV or chemical mutagenesis (Mattern, I.E., van Noort J.M., van den Berg, P., Archer, D. B., Roberts, I.N. and van den Hondel, C. A., Isolation and characterization of mutants of *Aspergillus niger* deficient in extracellular proteases. Mol Gen Genet. 1992 Aug;234(2):332-6.).

Preferably, the deficiency brought in the NHR pathway is an inducible one. This can be reached by replacing the endogenous regulatory regions of the gene encoding the component involved in NHR by new regulatory regions, preferably by using a repressible or regulatable promoter, more preferably by using a promoter that can be switch on/off: by glucose repression, or ammonia repression, or pH repression. Examples of glucose-repressed promoters are the *Penicillium chrysogenum pcbAB* promoter (Martin JF, Casqueiro J, Kosalkova K, Marcos AT, Gutierrez S. Penicillin and cephalosporin biosynthesis: mechanism of carbon catabolite regulation of penicillin production. Antonie Van Leeuwenhoek. 1999 Jan-Feb;75(1-2):21-31. Review.) or the *Aspergillus niger* glucoamylase promoter. Examples of on/off switchable promoters are described in the following publications:

- An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast: Belli et al, (1998) Nucl. Acid Research. vol 26, n.4:942-947,

- A light-switchable gene promoter system: Shimizu-Sato_et al, (2002) Nat. Biotech. Vol 20, no 10:1041-1044.

According to a preferred embodiment, the filamentous fungus is deficient in at least one of its endogenous genes, which are homologous with the following yeast genes involved in the NHR pathway KU70, KU80, RAD50, MRE11, XRS2 and SIR4 (van den Bosch et al (2002): DNA double-strand break repair by homologous recombination. Biol. Chem. Vol.383:873-892 and Allen et al, (2003): Interactive competition between homologous recombination and non-homologous end joining. Mol. Cancer Res., vol 1:913-920).

All kinds of mutants having at least one component involved in NHR, which is no longer capable or at least significantly less capable to perform its function in the process of NHR, are mutants contemplated by the present invention. Preferably, the component involved in NHR has been inhibited so that the efficiency of the NHR pathway in the obtained mutant is less than 90% of the activity in the parent cell it originates from under the same conditions as measured in the assay defined earlier, even preferably less than 85%, more preferably less than 80%, even more preferably less than 70%, most preferably less than 50%.

According to a preferred embodiment, the parent filamentous fungus is one of filamentous fungus as defined under the section host cell.

Preferably, the filamentous fungus cell is deficient in at least one of the following genes:

- *hdfA* as identified in SEQ ID NO: 2 or 19 or homologues thereof, or
- *hdfB* as identified in SEQ ID NO: 5 or 22 or homologues thereof, or

or both.

According to another preferred embodiment, the filamentous fungus has the amount of at least one of the proteins encoded by these genes *hdfA* and *hdfB* that is decreased upon induction.

According to another preferred embodiment, the down regulation of the expression level of at least one gene or DNA sequence may be obtained by genetic modification by submitting the filamentous fungus to mutagenesis. Filamentous fungal cells may be subjected to random mutagenesis and subsequently to a selection assay to isolate the mutants with improved HR from the whole diverse population of mutants.

According to a preferred embodiment of the present invention, one of the filamentous fungal cell defined under the section host cell is used as starting strain to perform the mutagenesis.

For example, the starting strain is subjected to UV irradiation so that the survival percentage is ranged between 0.001% and 60%. Preferably, the survival percentage is ranged between 0.01% and 50%. It is well known to the skilled person that conidiospores is the preferred material to mutagenize filamentous fungi by physical or chemical means. Mutants may however also be obtained from mycelium cells. Also, other mutagenic treatments than UV can be applied as chemical agents (e.g. NTG). The selection method described herein may be applied to select mutants obtained from either conidiospores or mycelium cells.

Preferably the mutagenesis is applied to conidiospores. UV irradiation is preferably applied for different times such as 7.5, 15 and 30 minutes to obtain mild, medium and strong mutation rate levels in the cells. The mutated samples may either be directly re-sporulated or incubated for an extended recovery period in a rich medium such as YNB or YEPD (see definition in example. 9) before sporulation was induced (for example as described in example 9).

The sporulated batches may be then tested for their efficiency in gene targeting. This could be tested by the following method. Protoplasts may be transformed with at least one, preferably two or more DNA fragments carrying expression cassettes of functional selection markers. The selection marker genes in the expression cassettes can be selected from a number of marker genes that are useful for transformation of filamentous fungi. By way of example these markers include but are not limited to dominant and bi-directional selection marker gene such as an acetamidase (*amdS*) gene (EP 635 574 B or WO 97/06261), auxotrophic marker genes such as *argB*, *trpC*, or *pyrG* and antibiotic resistance genes providing resistance against e.g. phleomycin, hygromycin Bor G418. Preferably the selection markers used are the *ble* and *amdS* genes. The *amdS* cassette used is the *A.nidulans* coding sequence fused to the *A.nidulans gpdA* promoter (EP635 574 B). *amdS* genes from other filamentous fungus may also be used (WO 97/06261). The gene *ble* encodes a protein capable of conferring resistance to phleomycin. The gene *amdS* encodes a protein enabling cells to grow on acetamide as the sole nitrogen source (as described in EP635 574B). Techniques applied for the transfer of DNA to protoplasts of filamentous fungi are well known in the art and are described in many references, including Finkelstein and Ball (eds.), Biotechnology of filamentous fungi,

technology and products, Butterworth-Heinemann (1992); Bennett and Lasure (eds.) More Gene Manipulations in fungi, Academic Press (1991); Turner, in: Pühler (ed), Biotechnology, second completely revised edition, VHC (1992). The Ca-PEG mediated protoplast transformation is used as described in EP635574B.

To select targeted integration of these two expression cassettes to two distinct specific loci in the filamentous fungi genome short homologous stretches of DNA may be added for example via PCR on both sides of the DNA fragments. Several types of construct could be made to improve the chances to select a mutant having an improved targeting efficiency: the homologous stretches of DNA could typically vary from 30bp to 1000bp, preferably 30bp to 700bp, more preferably 30bp to 500bp, even more preferably 30bp to 300bp, more preferably preferably 30bp to 200bp, even more preferably 30bp to 100 bp and most preferably 30bp. In theory all loci in the filamentous fungi genome could be chosen for targeting integration of the expression cassettes. Preferably, the locus wherein targeting will take place is such that when the wild type gene present at this locus has been replaced by the gene comprised in the expression cassette, the obtained mutant will display a change detectable by a given assay. Preferably the locus is the *niaD* locus, thereby disrupting the nitrate reductase gene (Gouka RJ, van Hartingsveldt W, Bovenberg RA, van den Hondel CA, van Gorcom RF. Cloning of the nitrate-nitrite reductase gene cluster of *Penicillium chrysogenum* and use of the *niaD* gene as a homologous selection marker. J Biotechnol. 1991 Sep;20(2):189-99), enabling direct selection of transformants on plates containing chlorate, as cells become resistant to chlorate. Another preferred locus is the *sutB* locus, thereby disrupting the sulphate permease gene (van de Kamp M, Pizzinini E, Vos A, van der Lende TR, Schuurs TA, Newbert RW, Turner G, Konings WN, Driessen AJ. Sulfate transport in *Penicillium chrysogenum*: cloning and characterization of the *sutA* and *sutB* genes. J. Bacteriol. 1999 Dec;181(23):7228-34), enabling direct selection of transformants on plates containing selenate. Mutants with both selection markers present and having the two alterations resulting from the inactivation of the genes present at the integration loci are strains with improved targeted integration.

According to another preferred embodiment, the mutant filamentous fungus having a lowered efficiency in the NHR pathway, or a decreased NHR/HR ratio and/or an elevated efficiency of the HR pathway is obtained by decreasing, more preferably partially or most preferably completely inhibiting a component involved in NHR.

Partial or complete inhibition of a component involved in NHR can be obtained by different methods, for example by an antibody directed against such a component or a chemical inhibitor or a protein inhibitor or a physical inhibitor (Tour O. et al, (2003) Nat. Biotech: Genetically targeted chromophore-assisted light inactivation. Vol.21. no. 12:1505-1508) or peptide inhibitor or an anti-sense molecule or RNAi molecule (R.S. Kamath_et al, (2003) Nature: Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi.vol. 421, 231-237). Irrespective of the kind of (partial or more preferably complete) inhibition it is important that a component involved in NHR is no longer capable or at least significantly less capable to perform its function in the process of NHR as defined above.

Components involved in NHR comprise filamentous fungal homologues of yeast KU70, RAD50, MREII, XRS2, LIG4, SIR4, KU80, LIFL or NEIL or associating components. Because the nomenclature of genes differs between organisms a functional equivalent or a functional and/or a functional fragment thereof, all defined herein as being capable of performing (in function, not in amount) at least one function of the yeast genes KU70, RAD50, MREII, XRS2, LIG4, SIR4, KU80, LIFL or NEIL are also included in the present invention. By transiently (partially or more preferably completely) inhibiting a component involved in NHR a nucleic acid is integrated at any desired position without permanently modifying a component involved in NHR and preventing unwanted side effects caused by the permanent presence of such a modified component involved in NHR.
In addition of the above-mentioned techniques or as an alternative, it is also possible to obtain a lowered NHR efficiency by inhibiting the activity of proteins, which are involved in NHR or to re-localize the NHR involved proteins by means of alternative signal sequences (Ramon de Lucas, J., Martinez O, Perez P., Isabel Lopez, M., Valenciano, S. and Laborda, F. The *Aspergillus nidulans* carnitine carrier encoded by the *acuH* gene is exclusively located in the mitochondria. FEMS Microbiol Lett. 2001 Jul 24;201(2):193-8.) or retention signals (Derkx, P. M. andMadrid, S. M. The foldase CYPB is a component of the secretory pathway of *Aspergillus niger* and contains the endoplasmic reticulum retention signal HEEL. Mol. Genet. Genomics. 2001 Dec;266(4):537-45.).

Alternatively or in combination with above-mentioned techniques, inhibition of protein activity can also be obtained by UV or chemical mutagenesis (Mattern, I.E., van Noort J.M., van den Berg, P., Archer, D. B., Roberts, I.N. and van den Hondel, C. A., Isolation

and characterization of mutants of *Aspergillus niger* deficient in extracellular proteases. Mol Gen Genet. 1992 Aug;234(2):332-6.) or by the use of inhibitors like the proteasomal inhibitor of Affinity (clasto-lactacystin-β-lactone, Affinity Research Products Ltd., CW8405-Z02185.).

According to another preferred embodiment, the steering towards HR comprises adding an excess of small double stranded polynucleotides able to bind and thereby limit the expression of NHR components, next to the polynucleotide to be integrated (Agrawal N. et al,: RNA interference: biology, mechanism and applications. Microbiol. Mol. Biol. Rev., vol. 67, no. 4:657-685).

In a preferred embodiment the invention provides a method for increasing the efficiency of targeted integration of a polynucleotide to a pre-determined site, whereby said polynucleotide has homology at or around the said pre-determined site, in a filamentous fungus with a preference for NHR comprising steering an integration pathway towards HR by providing a filamentous fungal cell, wherein the efficiency of the HR pathway has been elevated compared to the one of the parent filamentous fungus it originates from under the same conditions. The efficiency of the HR pathway is preferably assayed by the same assay as the one used for determining the NHR/HR ratio. According to a preferred embodiment, the parent organism is one of the filamentous fungi as defined in the section host cell.
Elevated means at least higher than in the parental cell the obtained cell originates from. Preferably, elevated means twice higher, more preferably three times higher, even more preferably four times higher, most preferably more than four times higher using northern, or western blotting or array technique or a phenotypic screen.

According to another preferred embodiment, the filamentous fungus has the expression level of at least one gene involved in HR, which has been up regulated by comparison to the expression level of the same gene in the filamentous fungal cell it originates from. This can be achieved by increasing the expression level of a gene encoding a component involved in HR and/or by increasing the expression level of a component involved in HR and/or by (temporarily) increasing the activity of the component involved in HR.

Preferably, the filamentous fungus obtained has the expression of a gene involved in HR, which has been up regulated by comparison to the expression of said gene in the filamentous fungal cell it originates from.

The expression level of a DNA sequence is up regulated when the expression level of this specific DNA sequence in the obtained filamentous fungus is higher than the expression level of the same DNA sequence in the parental filamentous fungus it originates from, preferably three times higher, more preferably four times higher, most preferably more than four times higher using northern, or western blotting or array technique. According to a preferred embodiment, the parent organism is one of the filamentous fungi as defined in the section host cell.

The up regulation of the expression level of at least one DNA sequence may be obtained by genetic manipulation by one of the following techniques or by a combination thereof:

c. using recombinant genetic manipulation techniques,

d. submitting the filamentous fungus to mutagenesis,

Alternatively or in combination with above-mentioned techniques and according to another preferred embodiment, the up regulation of the expression level of at least one gene or DNA sequence may be obtained by submitting the filamentous fungus to an activating compound/composition.

The filamentous fungus may be subsequently selected by monitoring the expression level of said DNA sequence and optionally the efficiency of the HR pathway of the filamentous fungus. The HR efficiency of a filamentous fungus may be measured by the efficiency of the targeted integration of a given polynucleotide sequence into a pre-determined site in the genome of the filamentous fungus using given homology region(s).

Preferably, the up regulation of the expression of at least one DNA sequence is made with recombinant genetic manipulation techniques such as defined in step a. to obtain a recombinant filamentous fungus. Preferably step a. comprises transforming the filamentous fungus with a DNA construct comprising the DNA sequence, preferably said DNA sequence being operationally linked to a promoter of a highly expressed gene. The chosen promoter may be stronger than the endogenous promoter of the DNA sequence

to be over expressed. The promoter for expression of the DNA sequence is preferably derived from a highly expressed fungal gene.

A number of preferred highly expressed fungal genes are given by way of example: the amylase, glucoamylase, alcohol dehydrogenase, xylanase, glyceraldehyde-phosphate dehydrogenase or cellobiohydrolase genes from *Aspergilli* or *Trichoderma*. Most preferred highly expressed genes for these purposes are an *Aspergillus niger* glucoamylase gene, an *Aspergillus oryzae* TAKA-amylase gene, an *Aspergillus nidulans gpdA* gene or a *Trichoderma reesei* cellobiohydrolase gene. A glucoamylase promoter is the most preferred promoter to be used. These highly expressed genes are suitable both as target loci for integration of cloning vectors and as source of highly expressed fungal genes.

According to another preferred embodiment, step a. comprises increasing the copy number of the DNA sequence into the filamentous fungal cell, preferably by integrating into its genome copies of the DNA sequence, more preferably by targeting the integration of the DNA sequence at a highly expressed locus, preferably at a glucoamylase locus.

The up regulation of the expression of the DNA sequence may be reached by increasing the copy number of the DNA sequence by introducing at least one copy of the DNA sequence into the filamentous fungus or by changing for a stronger promoter or changing for a gene encoding a protein with better kinetics and/or lifetime. The DNA sequence may be present on a plasmid or integrated into the genome. The skilled person can choose amongst two alternative possibilities:

- over express at least one endogenous DNA sequence of the filamentous fungus being involved in the HR pathway. In this case, the filamentous fungus comprises several copies of its endogenous DNA sequence.
- over express at least one heterologous DNA involved in HR. In this case, the filamentous fungus would have its endogenous DNA sequence involved in HR and, in addition at least one copy of a heterologous DNA sequence involved in HR.This heterologous DNA sequence is an homologue of its corresponding endogenous DNA sequence.

The filamentous fungus can be transformed with one or more copy of the DNA sequence (derived from *inter alia* Tilburn et al, 1983, Gene, 26:205-221). The DNA sequence can be either stably integrated into the genome of the filamentous fungus or introduced into the cell as part of a DNA molecule capable of autonomous replication. The DNA

sequence is preferably present on a cloning vector. Any cloning vector capable of transforming a filamentous fungal host cell is suitable for use in the present invention. Cloning vectors for use in the invention thus comprise integrative cloning vectors, which integrate at random or at a predetermined target locus in the chromosomes of the filamentous fungal host cell, as well as autonomously maintained cloning vectors such as vectors comprising the AMA1-sequence. In a preferred embodiment of the invention, the integrative cloning vector comprises a DNA fragment, which is homologous to a DNA sequence in a predetermined target locus in the genome of the filamentous fungal host cell for targeting the integration of the cloning vector to this predetermined locus. In order to promote targeted integration, the cloning vector is preferably linearized prior to transformation of the host cell. Linearization is preferably performed such that at least one but preferably either end of the cloning vector is flanked by sequences homologous to the target locus. The length of the homologous sequences flanking the target locus is preferably at least 30bp, preferably at least 50 bp, preferably at least 0.1kb, even preferably at least 0.2kb, more preferably at least 0.5 kb, even more preferably at least 1 kb, most preferably at least 2 kb.

Preferably, the DNA sequence in the cloning vector, which is homologous to the target locus is derived from a highly expressed locus meaning that it is derived from a gene, which is capable of high expression level in the filamentous fungal host cell. A gene capable of high expression level, i.e. a highly expressed gene, is herein defined as a gene whose mRNA can make up at least 0.5% (w/w) of the total cellular mRNA, e.g. under induced conditions, or alternatively, a gene whose gene product can make up at least 1% (w/w) of the total cellular protein, or, in case of a secreted gene product, can be secreted to a level of at least 0.1 g/l (as described in EP 357 127 B1).

To increase even more the number of copies of the DNA sequence to be over expressed the technique of gene conversion as described in WO98/46772 may be used.

The skilled person will appreciate the possibility that the homologous DNA sequence for targeting and the promoter sequence can coincide in one DNA fragment. The list of highly expressed genes given above is also suited as target locus.

An example of an autonomously maintained cloning vector is a cloning vector comprising the AMA1-sequence. AMA1 is a 6.0-kb genomic DNA fragment isolated from *Aspergillus nidulans*, which is capable of Autonomous Maintenance in *Aspergillus* (see e.g. Aleksenko and Clutterbuck (1997), Fungal Genet. Biol. **21**: 373-397).

According to another preferred embodiment of the method of the invention, step a. comprises transforming the filamentous fungus with a DNA construct comprising a selection marker gene. The selection marker gene in the cloning vector can be selected from a number of marker genes that are useful for transformation of filamentous fungi. By way of example these markers include but are not limited to dominant and bi-directional selection marker genes such as an *amdS* gene (EP 635574, WO 97/06261), auxotrophic marker genes such as *argB*, *trpC*, or *pyrG* and antibiotic resistance genes providing resistance against e.g. phleomycin, hygromycin B or G418.The use of a dominant and bi-directional selection marker gene is preferred. Preferably an *amdS* gene is preferred, more preferably an *amdS* gene from *Aspergillus nidulans* or *Aspergillus niger*. A most preferred selection marker gene is the *A.nidulans amdS* coding sequence fused to the *A.nidulans gpdA* promoter (see EP635574). *AmdS* genes from other filamentous fungus may also be used (WO 97/06261). The *amdS* selection marker gene has the advantage it can be used several times in the same strain to introduce, over express and/or delete distinct DNA sequences. By means of counterselection on fluoracetamide media as described in EP 635574, the resulting strain is marker free and can be used for further gene modifications.

Alternatively or in addition with above-mentioned techniques, up regulation of the expression of a DNA sequence can be reached using UV or chemical mutagenesis (Mattern, I.E., van Noort J.M., van den Berg, P., Archer, D. B., Roberts, I.N. and van den Hondel, C. A., Isolation and characterization of mutants of *Aspergillus niger* deficient in extracellular proteases. Mol Gen Genet. 1992 Aug;234(2):332-6.).

In addition and/or in combination with up regulation of expression of DNA sequences involved in HR, it is also possible to obtain an increased HR efficiency by increasing the activity of proteins involved in HR by UV or chemical mutagenesis (Mattern, I.E., van Noort J.M., van den Berg, P., Archer, D. B., Roberts, I.N. and van den Hondel, C. A., Isolation and characterization of mutants of *Aspergillus niger* deficient in extracellular proteases. Mol Gen Genet. 1992 Aug;234(2):332-6.).

The skilled person would understand that to achieve the up regulation of the expression of a DNA sequence, one may use each of the described technique either separately or in combination.

The skilled person would also understand that to obtain a filamentous fungus with an increased HR/NHR ratio, and/or with a lowered NHR efficiency and/or an elevated HR efficiency, one may use at least one of each technique described for respectively down and up regulating the expression of a given gene in a filamentous fungus. Preferably, all the techniques performed on the filamentous fungus to obtain a recombinant filamentous fungus having both a lowered NHR efficiency and an elevated HR efficiency have been performed using a dominant and bi-directional selection marker, preferably an *amdS* gene more preferably an *amdS* gene from *Aspergillus nidulans* or *Aspergillus niger.*
The obtained filamentous fungus may be subsequently selected by monitoring the expression level of said DNA sequence as described before by using for example northern and/or western blotting and/or array and/or phenotype screening. Optionally, the efficiency of the NHR and/or HR pathways of the cell is monitored. The efficiency of these pathways of a filamentous fungus may be monitored as defined earlier on.

Preferably, the modification brought in the HR pathway is an inducible one. This can be reached by replacing the endogenous regulatory regions of the gene encoding the component involved in HR by inducible regulatory regions, preferably by using an inducible promoter. Examples of inducible promoters are the glucoamylase promoter of *Aspergillus niger*, the TAKA amylase promoter of *Aspergillus oryzae*, the *paf* promoter (Marx,F., Haas,H., Reindl,M., Stoffler,G., Lottspeich,F. and Redl,B. Cloning, structural organization and regulation of expression of the *Penicillium chrysogenum paf* gene encoding an abundantly secreted protein with antifungal activity Gene 167 (1-2), 167-171 (1995) or the *pcbC* promoter of *Penicillium chrysogenum* (Martin JF, Casqueiro J, Kosalkova K, Marcos AT, Gutierrez S. Penicillin and cephalosporin biosynthesis: mechanism of carbon catabolite regulation of penicillin production. Antonie Van Leeuwenhoek. 1999 Jan-Feb;75(1-2):21-31. Review.) or the switch on/off systems earlier cited for down regulation of the expression of genes involved in NHR.

According to a preferred embodiment, the genes involved in the HR pathway, which are modified are the following genes or homologues thereof: RAD51, RAD52.

All kinds of mutants having at least one component involved in HR, which is more capable or at least significantly more capable to perform its function in the process of HR

are mutants contemplated by the present invention. Preferably, the activity of the components involved in HR has been modified so that the efficiency of the HR pathway is more than 110% of the efficiency in the parent cell it originates from under the same conditions as measured in the assay defined earlier, more preferably more than 200%, most preferably more than 500%. According to a preferred embodiment, the parent organism is one of the filamentous fungi as defined under the section host cell. Methods according to the present invention, as extensively but not limiting discussed above, can be used in a wide variety of applications. Some specific applications are described below.

Host cell

Accordingly, the present invention further relates to the filamentous fungus per se, which is preferably used in the method of the invention for increasing the efficiency of targeted integration of a polynucleotide to a pre-determined site into the genome of said filamentous fungal cell, said filamentous fungus having a preference for NHR, and wherein said polynucleotide has a region of homology with said pre-determined site and said method comprising steering an integration pathway towards HR. The characteristics of the filamentous fungus that can be used in this method have been earlier defined.

The filamentous fungus preferably used in the method of the invention is a mutant originating from a parent cell, wherein the ratio of NHR/HR is decreased and/or wherein the efficiency of the NHR pathway has been lowered and/or the efficiency of the HR pathway has been elevated in said mutant cell as compared to said ratio and said efficiencies in said parent organism under the same conditions. The assay used to determine the ratio NHR/HR and/or the efficiency of the NHR pathway and/or the efficiency of the HR pathway has been earlier described .

The host cell of the present invention is a filamentous fungus, which is capable of being transformed with a cloning vector. For most filamentous fungi tested thus far it was found that they could be transformed using transformation protocols developed for *Aspergillus* (derived from *inter alia* Tilburn *et al.* 1983, Gene 26 :205-221). The skilled person will recognise that successful transformation of the filamentous fungal host species is not limited to the use of vectors, selection marker systems, promoters and transformation protocols specifically exemplified herein.

A filamentous fungus is herein defined as a eukaryotic microorganism of the subdivision Eumycotina in filamentous form, i.e. the vegetative growth of which occurs by hyphal

elongation. Preferred filamentous fungal host cells are selected from the group consisting of the genera *Aspergillus*, *Trichoderma*, *Fusarium*, *Penicillium*, and *Acremonium*.

In a more preferred embodiment of the invention, the filamentous fungal host cell is selected from the group consisting of *A.nidulans*, *A.oryzae*, *A.sojae*, *Aspergilli* of the *A.niger* Group, *Trichoderma reesei* and *Penicillium species*. Preferably the *Penicillium* is a *Penicillium chrysogenum* or *Penicillium citrinum* species.

The *A.niger* group is herein defined according to Raper and Fennell (1965, In: The Genus *Aspergillus*, The Williams & Wilkins Company, Baltimore, pp 293-344) and comprises all (black) *Aspergilli* therein included by these authors. Most preferred filamentous fungal host cells are selected from the group consisting of *Aspergilli* of the *A.niger* group, *A.oryzae*, *Trichoderma reesei* and *Penicillium chrysogenum*.

According to a preferred embodiment, the parent organism is the deposited filamentous fungus cell *Aspergillus niger* CBS 513.88, *Aspergillus oryzae* ATCC 20423, IFO 4177, ATCC 1011, ATCC 9576, ATCC14488-14491, ATCC 11601, ATCC12892, *Penicillium chrysogenum* CBS 455.95 or or *Penicillium citrinum* ATCC 38065, *Penicillium chrysogenum* P2, *Acremonium* chrysogenum ATCC 36225 or ATCC 48272, *Trichoderma reesei* ATCC 26921 or ATCC 56765 or ATCC 26921, *Aspergillus sojae* ATCC11906, *Chrysosporium lucknowense* ATCC44006, *Claviceps paspali* CBS110.22, *Claviceps purpurea* CBS164.59, *Penicillium brevicompactum* ATCC 9056, *Aspergillus terreus* ATCC 20542, *Aspergillus nidulans* ATCC 28901and or derivatives thereof.

According to another preferred embodiment, the filamentous fungal cell of the invention has a ratio NHR/HR, which is at least 200, at least 50, at least 10 as measured by the following assay. Preferably the ratio of the filamentous fungal cell is at least 1, more preferably at least 0.5, even more preferably at least 0.1, even more preferably at least 0.05, even more preferably at least 0.01 even more preferably at least 0.005 even more preferably at least 0.001 even more preferably at least 0.0005 even more preferably at least 0.0001 and most preferably at least 0.00001.

According to a more preferred embodiment, the filamentous fungal cell of the invention has a ratio NHR/HR, which is less than 200, even more preferably less than 50, less than 10 as measured by the following assay. Even more preferably the ratio of the filamentous fungal cell is less than 1, even more preferably less than 0.5, even more

preferably less than 0.1, even more preferably less than 0.05, even more preferably less than 0.01 even more preferably less than 0.005 even more preferably less than 0.001 even more preferably less than 0.0005 even more preferably less than 0.0001 and most preferably less than 0.00001.

The ratio of NHR/HR is preferably measured by the assay as described in WO 02/052026 (table 2, p23).

Preferably, the filamentous fungal cell is deficient in a gene encoding a component involved in NHR, and/or has a decreased level of a component involved in NHR.
Even more preferably, the filamentous fungal cell is deficient in at least one of the following genes: *hdfA* or homologues thereof as identified in SEQ ID NO: 2 or 19, *hdfB* or homologues thereof as identified in SEQ ID NO: 5, or 22 or both, and/or has, preferably a decreased amount of at least one of the proteins encoded by these genes.
Most preferably, the filamentous fungal cell is inducibly deficient in at least one of the following genes: *hdfA* or homologues thereof as identified in SEQ ID NO: 2 or 19, *hdfB* or homologues thereof as identified in SEQ ID NO: 5, or 22 or both, and/or has, preferably inducibly, a decreased amount of at least one of the proteins encoded by these genes.
According to another preferred embodiment, the filamentous fungal cell is such that in its genome, a gene involved in NHR has been replaced by a non-functional gene or by a selection marker or by another gene.
According to another preferred embodiment, the mutant has an increased level of a component involved in HR.
The filamentous fungus according to the invention may have been obtained by molecular biology techniques. A filamentous fungus obtained by such a genetic engineering approach is defined as a recombinant filamentous fungus. However, a recombinant filamentous fungus in the context of the invention could have been subjected earlier in time to mutagenesis technique to reach another wanted effect. According to a most preferred embodiment, the filamentous fungus obtained is a recombinant filamentous fungus.

Use of the host cell of the invention

According to a preferred embodiment, there is provided a method which comprises at least the steps of introducing a polynucleotide of interest into the filamentous fungus of the invention, for example by the process of transformation or electroporation, and integration of said polynucleotide in the genetic material of said cell. Integration is a complex process wherein a nucleic acid sequence becomes part of the genetic material of a host cell. One step in the process of nucleic acid integration is recombination; via recombination nucleic acid sequences are exchanged or inserted and the introduced nucleic acid becomes part of the genetic material of a host cell. In principle two different ways of recombination are possible: homologous and illegitimate or NHR. Most (higher) eukaryotes do not or at least not significantly practice HR although the essential proteins to accomplish such a process are available. One reason for this phenomenon is that frequent use of homologous recombination in (higher) eukaryotes could lead to undesirable chromosomal rearrangements due to the presence of repetitive nucleic acid sequences. To accomplish HR via a method according to the invention, it is important to provide a polynucleotide, which has homology with a pre-determined site. It is clear to a person skilled in the art that the percentage of homology and the length of (a) homologous region(s) play(s) an important role in the process of homologous recombination. The percentage of homology is preferably close to 100%. A person skilled in the art is aware of the fact that lower percentage of homology are also used in the field of homologous recombination, but dependent on, for example, the regions of homology and their overall distribution, can lead to a lower efficiency of HR but are still useful and therefore included in the present invention. Furthermore, the length of a (nearly) homologous region is approximately 3 kb which is sufficient to direct homologous recombination. At least one homologous region is necessary for recombination but more preferably two homologous regions flanking the nucleic acid of interest are used for targeted integration. The researcher skilled in the art knows how to select the proper percentage of homology, the length of homology and the amount of homologous regions. By providing such a homology a nucleic acid is integrated at every desired position within the genetic material of a host cell. It is clear to a person skilled in the art that the invention as disclosed herein is used to direct any nucleic acid (preferably DNA) to any pre-determined site as long as the length of homology and percentage of homology are high enough to provide/enable HR.

Before the present invention was made, a polynucleotide could not have always easily been integrated at every desired position into the genome of a given filamentous fungus. The method according to the invention is applied, for example, to affect the gene function in various ways, not only for complete inactivation but also to mediate changes in the expression level or in the regulation of expression, changes in protein activity or the subcellular targeting of an encoded protein. Complete inactivation, which can usually not be accomplished by existing methods such as antisense technology or RNAi technology (Zrenner R, Willmitzer L, Sonnewald U. Analysis of the expression of potato uridinediphosphate-glucose pyrophosphorylase and its inhibition by antisense RNA. Planta. (1993);190(2):247-52.) is useful for instance for the inactivation of genes controlling undesired side branches of metabolic pathways, for instance to increase the production of specific secondary metabolites such as (beta-lactam) antibiotics or carotenoids. Complete inactivation is also useful to reduce the production of toxic or unwanted compounds (chrysogenin in *Penicillium*; Aflatoxin in *Aspergillus:* MacDonald KD et al,: heterokaryon studies and the genetic control of penicillin and chrysogenin production in *Penicillium chrysogenum*. J Gen Microbiol. (1963) 33:375-83). Complete inactivation is also useful to alter the morphology of the organism in such a way that the fermentation process and down stream processing is improved.

The invention allows to replace existing regulatory sequences by alternative regulatory sequences to alter expression of endogenous genes (e. g. expression in response to specific inducers.

One aspect of the present invention relates to the replacement of an active gene by an inactive gene according to a method of the invention. Complete inactivation, which can usually not be accomplished by existing methods such as antisense technology or RNAi technology, is useful for instance for the inactivation of genes controlling undesired side branches of metabolic pathways, for instance to increase the quality of bulk products such as starch, or to increase the production of specific secondary metabolites or to inhibit formation of unwanted metabolites.

Another aspect of the invention relates to the extensive metabolic reprogramming or engineering of a filamentous fungal cell. Introduction of complete new pathways and/or

modification of unwanted pathways will lead to the obtention of a cell specifically adapted for the production of a specific compound such as a protein or a metabolite.

Another aspect of the present invention relates to the replacement of an inactive or altered gene by an active gene. For example, after successive rounds of classical mutagenesis, it often occurs the selected filamentous fungal strain has some endogenous genes altered or even inactivated during the random mutagenesis process.

In yet another aspect of the invention there is provided a method to introduce a substance conferring resistance for an antibiotic substance to a filamentous fungal cell. In yet a further aspect of the invention, there is provided a method to confer a desired property to a filamentous fungal cell. In a preferred embodiment a gene delivery vehicle is used to deliver a desired polynucleotide to a predetermined site. Gene delivery vehicles are well known in the art and have been earlier described in the description.

Also another preferred method according to a further aspect of the invention is to effectuate predictable expression of transgenes encoding novel products, for example by replacing existing coding sequences of genes giving a desired expression profile by those for a desired novel product. According to a more preferred embodiment, the filamentous fungus provided by the invention further comprises a DNA construct comprising a desired gene coding for a desired protein to be produced.
Preferably, the desired gene encoding the desired protein to be produced is inserted into an expression vector, which is subsequently used to transform the obtained host cell. In the expression vector, the DNA sequence may be operationally linked to appropriate expression signals, such as a promoter, optionally a signal sequence and a terminator, which are capable of directing the expression and synthesis of the protein in the host organism.
More preferably, the desired gene is operationally linked to a promoter and to a secretion signal. The strategy, which can be used to express the desired gene is the same as the one described under the section up regulation of the expression of a DNA sequence, whose expression product is involved in HR: increasing copy number, targeting integration, use of a promoter of a highly expressed gene, choice of the selection marker gene and combinations thereof.

The desired protein is preferably an enzyme. If the protein is not naturally secreted, the polynuclelotide encoding the protein may be modified to have a signal sequence in accordance with techniques known in the art. The proteins, which are secreted may be endogenous proteins which are expressed naturally, but can also be heterologous. Heterologous means that the gene encoded by the protein is not produced under native condition in the wild type filamentous fungus. Examples of enzymes which may be produced by the filamentous fungi of the invention are carbohydrases, e.g. cellulases such as endoglucanases, β-glucanases, cellobiohydrolases or β-glucosidases, hemicellulases or pectinolytic enzymes such as xylanases, xylosidases, mannanases, galactanases, galactosidases, rhamnogalacturonases, arabanases, galacturonases, lyases, or amylolytic enzymes; phosphatases such as phytases, esterases such as lipases, proteolytic enzymes, oxidoreductases such as oxidases, transferases, or isomerases. More preferably, the desired gene encodes a phytase.

As another example existing coding sequences are modified so that the protein encoded has optimized characteristics for instance to make a protein with improved thermal characteristics and/or improved kinetic properties (Km, Kcat), and/or improved enzyme stability, and/or extended substrate range, and/or increased life span, etc.

The invention further relates to the use of the filamentous fungus of the invention for producing a polypeptide of interest. Alternatively, the filamentous fungus obtained may be used for producing a secondary metabolite. Preferred secondary metabolites are carotenoid compounds, beta-lactam compounds, drugs, anti-tumor compounds, etc.

Preferably, the filamentous fungus as obtained in the present invention is used for producing the desired protein by culturing the transformed host cell under conditions conducive to the expression of the DNA sequence encoding the desired protein, and recovering the desired protein as described for example in the following references:

- Li, Z. J., Shukla, V., Fordyce, A. P., Pedersen, A. G., Wenger, K. S., Marten, M. R. Fungal morphology and fragmentation behavior in a fed-batch Aspergillus oryzae fermentation at the production scale.
  Biotechnol Bioeng. 2000 Nov 5;70(3):300-12

- Withers, J. M., Swift, R. J., Wiebe, M. G., Robson, G. D., Punt, P. J., van den Hondel, C. A. Optimization and stability of glucoamylase production by recombinant strains of Aspergillus niger in chemostat culture. Biotechnol Bioeng. 1998 Aug 20;59(4):407-18.

- Amanullah, A., Christensen, L. H., Hansen, K., Nienow, A. W., Thomas, R. C. Dependence of morphology on agitation intensity in fed-batch cultures of Aspergillus oryzae and its implications for recombinant protein production. Biotechnol Bioeng. 2002 Mar 30;77(7):815-26.

DNA sequences and polypeptides encoded by these DNA sequences

According to a further aspect of the invention, there are provided the following isolated cDNA sequences:

SEQ ID NO: 2 *hdfA* from *A. niger*,

SEQ ID NO: 19 *hdfA* from *Penicillium chrysogenum*

SEQ ID NO: 5 *hdfB* from *A. niger*

SEQ ID NO: 22 *hdfB* from *Penicillium chrysogenum*

and homologues thereof.

Each SEQ ID NO: 1, 18, 4 and 21 corresponds respectively to the genomic DNA sequence associated with each cDNA sequence given above.

Each SEQ ID NO: 3, 20, 6 and 23 corresponds respectively to the protein sequence encoded by the respective cDNA sequence given above.

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein can be readily used to isolate the complete gene from filamentous fungi, in particular *A. niger* or *Penicillium chrysogenum* which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any

nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct for such errors.

"Homologous" is below defined. Homologous can be understood as meaning derived from other filamentous fungus than *Aspergillus niger* or *Penicillium chrysogenum*.
Full length DNA from other organisms can be obtained in a typical approach, using cDNA or genomic DNA libraries constructed from other organisms, e.g. filamentous fungi, in particular from the species *Aspergillus* or *Penicillium* by screening them.
The invention also encompasses paralogues of *hdfA* and/or *hdfB*. In the context of the invention, paralogues means DNA sequences homologous to SEQ ID NO: 1 or SEQ ID NO: 4 or SEQ ID NO: 18 or SEQ ID NO: 21 and derived from *A. niger* or *Penicillium chrysogenum* respectively.
For example, *Aspergillus* or *Penicillium* strains can be screened for homologous *hdfA* and/or *hdfB* polynucleotides by Northern blot analysis. Upon detection of transcripts homologous to polynucleotides according to the invention, cDNA libraries can be constructed from RNA isolated from the appropriate strain, utilizing standard techniques well known to those of skill in the art. Alternatively, a total genomic DNA library can be screened using a probe hybridisable to an *hdfA* and/or *hdfB* polynucleotide according to the invention.

Homologous gene sequences can be isolated, for example, by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of nucleotide sequences as taught herein.

The template for the reaction can be cDNA obtained by reverse transcription of mRNA prepared from strains known or suspected to express a polynucleotide according to the invention. The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a new *hdfA* and/or *hdfB* nucleic acid sequence, or a functional equivalent thereof.

The PCR fragment can then be used to isolate a full-length cDNA clone by a variety of known methods. For example, the amplified fragment can be labeled and used to screen a bacteriophage or cosmid cDNA library. Alternatively, the labeled fragment can be used to screen a genomic library.

PCR technology also can be used to isolate full-length cDNA sequences from other organisms. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis.

The resulting RNA/DNA hybrid can then be "tailed" (e.g., with guanines) using a standard terminal transferase reaction, the hybrid can be digested with RNase H, and second strand synthesis can then be primed (e.g., with a poly-C primer). Thus, cDNA sequences upstream of the amplified fragment can easily be isolated. For a review of useful cloning strategies, see e.g. Sambrook et al., *vide supra*; and Ausubel et al., *vide infra*.

"Homologous" can also be understood as meaning functional equivalents.

The terms "functional equivalents" and "functional variants" are used interchangeably herein. Functional equivalents of *hdfA* and/or *hdfB* DNA are isolated DNA fragments that encode a polypeptide that exhibits a particular function of the *hdfA* and/or *hdfB*. A functional equivalent of an *hdfA* and/or *hdfB* polypeptide according to the invention is a

polypeptide that exhibits at least one function as part of the NHR complex. Functional equivalents therefore also encompass biologically active fragments.

Functional protein or polypeptide equivalents may contain only conservative substitutions of one or more amino acids of sequences having SEQ ID NO: 3 or 6 or 20 or 23 or substitutions, insertions or deletions of non-essential amino acids. Accordingly, a non-essential amino acid is a residue that can be altered in one of these sequences without substantially altering the biological function. For example, amino acid residues that are conserved among the *hdfA* and/or *hdfB* proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, amino acids conserved among the *hdfA* and/or *hdfB* proteins according to the present invention are not likely to be amenable to alteration.

The term "conservative substitution" is intended to mean that a substitution in which the amino acid residue is replaced with an amino acid residue having a similar side chain. These families are known in the art and include amino acids with basic side chains (e.g.lysine, arginine and hystidine), acidic side chains (e.g. aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagines, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine tryptophan, histidine).

Functional nucleic acid equivalents may typically contain silent mutations or mutations that do not alter the biological function of encoded polypeptide. Accordingly, the invention provides nucleic acid molecules encoding *hdfA* and/or *hdfB* proteins that contain changes in amino acid residues that are not essential for a particular biological activity. Such *hdfA* and/or *hdfB* proteins differ in amino acid sequence from SEQ ID NO: 3 or 6, or 20 or 23 and yet retain at least one of their biological activities. In one embodiment the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises a substantially homologous amino acid sequence of at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequence shown in SEQ ID NO 3 or 6 or 20 or 23. For example, guidance concerning how to make phenotypically silent

amino acid substitutions is provided in Bowie, J.U. et al., Science 247:1306-1310 (1990) wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selects or screens to identify sequences that maintain functionality. As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require non-polar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie et al. and the references cited therein.

An isolated nucleic acid molecule encoding an *hdfA* and/or *hdfB* protein homologous to the protein according to SEQ ID NO: 3 or 6 or 20 or 23 can be created by introducing one or more nucleotide substitutions, additions or deletions into the coding nucleotide sequences according to SEQ ID NO: 2 or SEQ ID NO: 5, or SEQ ID NO: 19 or SEQ ID NO: 22 such that one or more amino acid substitutions, deletions or insertions are introduced into the encoded protein. Such mutations may be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis.

The term "functional equivalents" also encompasses orthologues of the *A. niger hdfA* and/or *hdfB* protein. Orthologues of the *A. niger hdfA* and/or *hdfB* protein are proteins that can be isolated from other strains or species and possess a similar or identical biological activity. Such orthologues can readily be identified as comprising an amino acid sequence that is substantially homologous to SEQ ID NO: 3 or 6 or 20 or 23.

"Homologous" can also be understood as meaning "substantially homologous".
The term "substantially homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with similar side chain) amino acids or nucleotides to a second amino acid or nucleotide sequence such that the first and the second amino acid or nucleotide sequences have a common domain. For example, amino acid or nucleotide sequences which contain a common domain having about 45%, preferably about 50%, preferably about 60%,

preferably about 65%, more preferably about 70%, even more preferably about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity or more are defined herein as sufficiently identical.

Also, nucleic acids encoding other *hdfA* and/or *hdfB* family members, that have a nucleotide sequence that differs from SEQ ID NO: 2 or 5 or 19 or 22, are within the scope of the invention. Moreover, nucleic acids encoding *hdfA* and/or *hdfB* proteins from different species, which thus have a nucleotide sequence which differs from SEQ ID NO: 2 or 5 or 19 or 22.

Nucleic acid molecules corresponding to variants (e.g. natural allelic variants) and homologues of the *hdfA* and/or *hdfB* DNA of the invention can be isolated based on their homology to the *hdfA* and/or *hdfB* nucleic acids disclosed herein using the cDNAs disclosed herein or a suitable fragment thereof, as a hybridisation probe according to standard hybridisation techniques preferably under highly stringent hybridisation conditions.

"Stringency" of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to reanneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature, which can be used. As a result, it follows that higher relative temperatures would tend to make the reaction conditions more stringent, while lower temperatures less so.

For additional details and explanation of stringency of hybridization reactions, see Ausubel et al, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

"Stringent conditions" or "high stringency conditions", as defined herein, may be identified by those that: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride / 0.0015 M sodium citrate / 0.1% sodium dodecyl sulfate at 50°C; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1 % bovine serum albumin / 0.1 % Ficdl / 0.1 % polyvinylpyrrolidone / 50mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C ; or (3) employ 50% formamide, 5 x SSC (0.75 M

NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1 % sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 Rg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a high-stringency wash consisting of 0.1 x SSC containing EDTA at 55°C.

"Moderately stringent conditions" may be identified as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989, and include the use of washing solution and hybridization conditions (e. g., temperature, ionic strength and % SDS) less stringent that those described above. An example of moderately stringent conditions is overnight incubation at 37°C in a solution comprising: 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1 x SSC at about 37-50 C. The skilled artisan will recognize how to adjust the temperature, lonic strength, etc. as necessary to accommodate factors such as probe length and the like. or by using an algorithm suitable for determining sequence similarity.

Homologous (similar or identical) sequences can also be determined by using a "sequence comparison algorithm". Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l Acad. Sci. USA 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection. An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, et al., J. Mol. Biol. 215: 403-410 (1990).

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www. ncbi. nlm. nih. gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of

length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached.

The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (see, e. g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P (N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, an amino acid sequence is considered similar to a protein such as a protease if the smallest sum probability in a comparison of the test amino acid sequence to a protein such as a protease amino acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. Preferably the similarity is at least 40% homology to one of the DNA sequences having SEQ ID NO:2, 5, 19 and 22. More preferably the similarity is at least 50%, more preferably, at least 60%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%.

In addition to naturally occurring allelic variants of the *hdfA* and/or *hdfB* sequence, the skilled person will recognise that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO: 2 or 5 or 19 or 22, thereby leading to changes in the amino acid sequence of the *hdfA* and/or *hdfB* protein without substantially altering the function of the *hdfA* and/or *hdfB* protein.

In another aspect of the invention, deteriorated *hdfA* and/or *hdfB* proteins are provided. Deteriorated *hdfA* and/or *hdfB* proteins are proteins, wherein at least one biological activity is decreased. Such proteins may be obtained by randomly introducing mutations along all or part of the *hdfA* and/or *hdfB* coding sequence, such as by saturation mutagenesis, and the resulting mutants can be expressed recombinantly and screened for biological activity. For instance, the art provides for standard assays for measuring their enzymatic activity and thus deteriorated proteins may easily be selected. Preferably, the assay is the one described earlier on (see for example WO02/052026 page 23 or the phenotypic screening assay).

In a preferred embodiment, the *hdfA* and/or *hdfB* protein has an amino acid sequence according to SEQ ID NO: 3 or 6 or 20 or 23. In another embodiment, the *hdfA* and/or *hdfB* polypeptide is substantially homologous to the amino acid sequence according to SEQ ID NO: 3 or 6 or 20 or 23 and retains at least one biological activity of a polypeptide according to SEQ ID NO:3 or 6 or 20 or 23, yet differs in amino acid sequence due to natural variation or mutagenesis as described above.

In a further preferred embodiment, the *hdfA* and/or *hdfB* protein has an amino acid sequence encoded by an isolated nucleic acid fragment capable of hybridising to a nucleic acid according to SEQ ID NO: 2 or 5 or 19 or 22, preferably under highly stringent hybridisation conditions.

Accordingly, the *hdfA* and/or *hdfB* protein is a protein which comprises an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequence shown in SEQ ID NO: 3 or 6 or 20 or 23 and retains at least one functional activity of the polypeptide according to SEQ ID NO: 3 or 6 or 20 or 23.

Functional equivalents of a protein according to the invention can also be identified e.g. by screening combinatorial libraries of mutants, e.g. truncation mutants, of the protein of the invention for a given activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein

sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods that can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening a subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations of truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

In addition to the *hdfA* and/or *hdfB* gene sequences shown in SEQ ID NO: 2 and 5 and 19 and 22, it will be apparent for the person skilled in the art that DNA sequence polymorphisms that may lead to changes in the amino acid sequence of the *hdfA* and/or *hdfB* protein may exist within a given population. Such genetic polymorphisms may exist in cells from different populations or within a population due to natural allelic variation. Allelic variants may also include functional equivalents.

Fragments of a polynucleotide according to the invention may also comprise polynucleotides not encoding functional polypeptides. Such polynucleotides may function as probes or primers for a PCR reaction.

Nucleic acids according to the invention irrespective of whether they encode functional or non-functional polypeptides, can be used as hybridization probes or polymerase chain reaction (PCR) primers. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having an *hdfA* and/or *hdfB* activity include, inter alia, (1) isolating the gene encoding the *hdfA* and/or *hdfB* protein, or allelic variants thereof from a cDNA library e.g. from other organisms than *A. niger* or *Penicillium chrysogenum*; (2) in situ hybridization (e.g. FISH) to metaphase chromosomal spreads to provide precise chromosomal location of the *hdfA* and/or *hdfB* gene as described in Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988); (3) Northern blot analysis for detecting expression of *hdfA* and/or *hdfB* mRNA in specific tissues and/or cells and 4) probes and primers that can be used as a diagnostic tool to analyse the presence of a nucleic acid hybridisable to the *hdfA* and/or *hdfB* probe in a given biological (e.g. tissue) sample.

Also encompassed by the invention is a method of obtaining a functional equivalent of an *hdfA* and/or *hdfB* gene or cDNA. Such a method entails obtaining a labelled probe that includes an isolated nucleic acid which encodes all or a portion of the sequence according to SEQ ID NO: 2 or 5 or 19 or 22 or a variant thereof; screening a nucleic acid fragment library with the labelled probe under conditions that allow hybridisation of the probe to nucleic acid fragments in the library, thereby forming nucleic acid duplexes, and preparing a full-length gene sequence from the nucleic acid fragments in any labelled duplex to obtain a gene related to the *hdfA* and/or *hdfB* gene.

In one embodiment, an *hdfA* and/or *hdfB* nucleic acid of the invention is at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to a nucleic acid sequence shown in SEQ ID NO: 1, or 2, or 4 or 5 or 18, or 19, or 21, or 22.

In another preferred embodiment an *hdfA* and/or *hdfB* polypeptide of the invention is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the amino acid sequence shown in SEQ ID NO: 3 or 6 or 20 or 23.

The invention relates to DNA sequences having SEQ ID NO: 1, or 2, or 4, or 5, or 18, or 19, or 21, or 22 per se and to homologues thereof as defined above. DNA sequences related to these DNA sequences and obtained by degeneration of the genetic code are also part of the invention. DNA sequences related to DNA SEQ ID NO: 2, 5, 19, and 22 and obtained by hybridisation (see former paragraph) are also part of the invention. Isolated polypeptide encoded by these DNA sequences or homologues thereof as defined above are also part of the invention. Polypeptides *hdfA* and *hdfB* have a function involved in NHR. All these polypeptides can be used in the method of the invention to obtain filamentous fungi, which may have improved targeting efficiencies.

The invention will be illustrated in more detail in the following examples. Such examples are not intended to limit the scope of the invention.

EXAMPLES

Example 1: Identification of the *hdf*A and *hdf*B genes and construction of the deletion vectors.

Genomic DNA of *Aspergillus niger* strain CBS513.88 was sequenced and analyzed. Two genes with translated proteins annotated as homologues to KU70 and KU80, were identified and named *hdf*A and *hdf*B respectively. Sequences of the *hdf*A and *hdf*B loci, comprising the open reading frame (ORF) (with introns) and approximately 1000 bp 5' and 3' of the genes, are shown in sequence listings 1 and 4. Gene replacement vectors for *hdf*A and *hdf*B were designed according to known principles and constructed according to routine cloning procedures (see figures 1 and 2). In essence, these vectors comprise approximately 1000 bp flanking regions of the *hdf* ORFs for homologous

recombination at the predestined genomic loci. In addition, they contain the *A. nidulans* bi-directional *amd*S selection marker, in-between direct repeats. The general design of these deletion vectors were previously described in EP635574B and WO 98/46772.

Example 2: Inactivation of the *hdf*A gene in *Aspergillus niger*

Linear DNA of deletion vector pDEL-HDFA (figure 1) was isolated and used to transform *Aspergillus niger* CBS513.88 using method earlier described (Biotechnology of Filamentous fungi: Technology and Products. (1992) Reed Publishing (USA); Chapter 6: Transformation pages 113 to 156). This linear DNA can integrate into the genome at the *hdf*A locus, thus substituting the *hdf*A gene by the *amd*S gene as depicted in figure 3. Transformants were selected on acetamide media and colony purified according to standard procedures as described in EP635574B. Spores were plated on fluoro-acetamide media to select strains, which lost the *amd*S marker. Growing colonies were diagnosed by PCR for integration at the *hdf*A locus and candidate strains tested by Southern analyses for deletion of the *hdf*A gene. Deletion of the *hdf*A gene was detectable by ~ 2,2 kb size reduction of DNA fragments covering the entire locus and hybridized to appropriate probes. Approximately 8 strains showed a removal of the genomic *hdf*A gene from a pool of approximately 400 initial transformants.
Strain dHDFA was selected as a representative strain with the *hdf*A gene inactivated.

Example 3: Inactivation of the *hdf*B gene in *Aspergillus niger*

Linear DNA of deletion vector pDEL-HDFB (figure 2) was isolated and used to transform the *Aspergillus niger* strain CBS513.88. This linear DNA can integrate into the genome at the *hdf*B locus, thus substituting the *hdf*B gene for *amd*S (figure 4). The same technique of gene replacement was used as the one described in example 2. Transformants were selected on acetamide media and colony purified according to standard procedures. Spores were plated on fluoro-acetamide media to select strains, which lost the *amd*S marker (EP 635574B). Growing colonies were diagnosed by PCR for integration at the *hdf*B locus and candidate strains tested by Southern analyses for deletion of the *hdf*B gene. Deletion of the *hdf*B gene was detectable by ~ 2,6 kb size reduction of DNA fragments covering the entire locus and hybridized to appropriate probes. Approximately 7 strains showed a removal of the genomic *hdf*B gene from a pool of approximately 370 initial transformants.
Strain dHDFB was selected as a representative strain with the *hdf*B gene inactivated.

Example 4: Inactivation of the *hdf*A and *hdf*B genes in *Aspergillus niger*

Linear DNA of deletion vector pDEL-HDFB (figure 2) was isolated and used to transform strain dHDFA obtained in Example 2. This linear DNA can integrate into the genome at the *hdf*B locus, thus substituting the *hdf*B gene for *amd*S (figure 4). The technique of gene replacement used is the one described in example 2.Transformants were selected on acetamide media and colony purified according to standard procedures. Spores were plated on fluoro-acetamide media to select strains, which lost the *amd*S marker. Growing colonies were diagnosed by PCR for integration at the *hdf*B locus and candidate strains tested by Southern analyses for deletion of the *hdf*B gene. Deletion of the *hdf*B gene was detectable by ~ 2,6 kb size reduction of DNA fragments covering the entire locus and hybridized to appropriate probes. Approximately 15 strains showed a removal of the genomic *hdf*B gene from a pool of approximately 380 initial transformants.

Strain dHDFAB was selected as a representative strain with both the *hdf*A and *hdf*B genes inactivated.

Example 5: Improved targeting for single homologous recombination events.

One mechanism by which DNA may integrate into the genome of *Aspergillus niger* at a predestined locus is through a single homologous recombination. Homologous DNA aligns and integrates at the genomic position by recombination (see figure 5). Two vectors were used to test the targeting efficiency through a single homologous recombination of *Aspergillus niger* strains obtained in examples 2, 3, and 4. The two vectors comprise regions homologous to the glucoamylase (*glaA*) locus to direct recombination and resulting integration (figure 5).

The first vector designed for such homologous integration has already been earlier described in WO 02/45524 (pGBFIN11-EPO). This vector contains the gene coding for the proline specific endoprotease.

The second vector (pGBFIN11-PLA) contains the gene coding for phospholipase A1 (PLA1) from *A. oryzae*. The gene encoding this enzyme has already been published (Watanabe I, et al, Biosci. Biotechnol. Biochem. (1999), Vol 63, numero 5, pages 820-826). This gene was cloned into pGBFIN11 using the same technique as described in WO 02/045524 for the cloning of the proline specific endoprotease gene in pGBFIN11-EPO.

Strains CBS513.88, dHDFA, dHDFB and dHDFAB were transformed with either pGBFIN11-EPO or pGBFIN11-PLA plasmids according to transformation techniques earlier described in example 2. The results obtained were the same with both plasmids used. We found respectively, 5 %, 10%, 10% and 10%, of transformants with plasmids integrated at the target locus. Hence, we concluded that the inactivation of at least one *hdf*-gene in *Aspergillus niger* leads to a significant increase of the targeting efficiency of these strains through a single homologous recombination event.

Example 6: Improved targeting for double homologous recombination events at several different loci.

The targeting efficiency was further assessed by transformation of the dHDFA strain with deletion vectors designed for the inactivation of a number of amylase encoding genes from the genome. Gene-flanking regions were cloned essentially as described in Example 1, and the resulting vectors were linearised and used to transform protoplasts of CBS513.88 and the dHDFA strain. The targeting frequency was assessed by PCR analyses and activity-based plate assays indicative of the inactivation of the corresponding genes. The latter was done by propagating transformants on PDA plates supplemented with 0.4% agar and subsequent staining with an iodine/potassium iodine solution (Lugol, Sigma L 6146). As can be seen in Table 1 below, the targeting frequency, as judged by PCR analyses and/or activity-based plate assays indicative of the inactivation of the corresponding genes, was significantly improved over that observed with the CBS513.88 strain.

Table 1. Targeting frequencies of several deletion vectors in the dHDFA strain as compared with strain CBS513.88

Gene	SEQ ID NO:	Plasmid	Targeting (%)		Increase (fold)
			CBS513.88	dHDFA	
amyBI	9	pDEL-AMYBI	18	83	4.6
amyBII	12	pDEL-AMYBII	17	79	4.6
amyA	15	pDEL-AMYA	6	57	9.5

These findings provide further support for our conclusion that inactivation of at least one of the *hdf* genes in *Aspergillus niger* results in a significant increase of the targeting efficiency of vectors for integration through double homologous recombination.

Example 7: The effect of size reduction of the homologous flanking regions of the *amyBII* gene on targeting frequencies.

In a separate series of experiments the effect of flanking region length on the transformation efficiency and targeting frequency through double homologous recombination was further investigated. Protoplasts of strains CBS513.88 and dHDFA were transformed with PCR fragments encompassing the *A. nidulans amdS* marker flanked by *amyBII* flanking regions of variable length. The data shown in Table 2 clearly demonstrate that, in addition to enhanced overall transformation efficiencies, targeting of the integrative cassettes was much improved in the dHDFA strain.

Table 2. Transformation efficiency and targeting frequencies of *amyBII* PCR deletion cassettes of variable length in the dHDFA strain and strain CBS513.88

Length (kb)	Nr. of transformants		Targeting (%)	
	CBS513.88	dHDFA	CBS513.88	dHDFA
1.0	13	84	46	87 [b]
0.5	0	7	n.d.[a]	
0.25	0	1	n.d.[a]	

[a], not determined
[b], combined % for three variants tested

Example 8: Phenotype analysis and production of polypeptide.

No phenotypic differences were observed during growth of the dHDFA, dHDFB or dHDFAB strains on solid media or shake flasks. Strains dHDFA, dHDFB and dHDFAB transformed with plasmids pGBFIN11-EPO or pBGFIN-PLA all secreted active enzyme into the medium as determined according to the following procedures.

Solid media was the potato dextrose agar (PDA) medium (Difco, POTATO DEXTROSE AGAR, cultivation medium, catalogus. nr. 213400, year 1996-1997).

Shake flask experiments were performed in 100 ml of the medium as described in EP 635 574 B at 34°C and 170 rpm in an incubator shaker using a 500 ml baffled shake flask. After four days of fermentation, samples were taken to determine either the proline specific endoprotease activity or the phospholipase activity.

The proteolytic activity of the proline specific endoprotease was spectrophotometrically measured in time at pH 5 and about 37°C using Z-Gly(cine)-Pro(line)-pNA as a substrate. 1U proline specific endoprotease is defined as the amount of enzyme which converts 1 micromol Z-Gly(cine)-Pro(line)-pNA per min at pH 5 and at 37°C.

To determine phospholipase PLA1 activity from *Aspergillus niger* (PLA1) spectrophotometrically, an artificial substrate is used: 1,2-dithiodioctanoyl phophatidylcholine (diC8, substrate). PLA1 hydrolyses the sulphide bond at the A1 position, dissociating thio-octanoïc acid. Thio-octanoïc acid reacts with 4,4 dithiopyridine (color reagent, 4DTDP), forming 4-thiopyridone. 4-Thiopyridone is in tautomeric equilibrium with 4-mercaptopyridine, which absorbs radiation having a wavelength of 334 nm. The extinction change at that wavelength is measured. One unit is the amount of enzyme that liberates of 1 nmol thio-octanoïc acid from 1,2-dithiodioctanoyl phosphatidylcholine per minute at 37°C and pH 4.0.

The substrate solution is prepared by dissolving 1 g diC8 crystals per 66 ml ethanol and add 264 ml acetate buffer. The acetate buffer comprises 0.1 M Acetate buffer pH 3.85 containing 0.2% Triton-X100. The colour reagent is a 11 mM 4,4-dithiodipyridine solution. It was prepared by weighting 5,0 mg 4,4-dithiodipyridine in a 2 ml eppendorf sample cup and dissolving in 1.00 ml ethanol. 1.00 ml of milli-Q water was added.

Interestingly, morphologic changes such as color differences or colony appearance occurred less frequent for transformants obtained from the dHDFA, dHDFB and dHDFAB strains than for transformants obtained from CBS513.88. This could be due to reduction of random integrations (NHR) thus preventing unexpected phenotypic changes.

Example 9: Isolation of *Penicillium* mutants with improved efficiency for homologous recombination by mutagenesis

To isolate mutants with an improved efficiency of gene targeting a combination of classical mutagenesis and molecular biology was applied. *Penicillium chrysogenum* (CBS 455.95) spores were obtained from colonies sporulating in YEPD (2% Yeast extract from Difco, 1% pepton from Difco, 2% glucose). These spores were washed in sterile tap water and 10 ml of a suspension containing 10^8 conidiospores per ml was subjected to UV irradiation at 254 nm (Sylvania, 15 Watts Black Light Blue tube, model

FT15T8/BLB). UV irradiation was applied for 7.5, 15 or 30 minutes while the suspensions were slowly shaken. These different irradation times were chosen to obtain mild, medium and strong mutation rate levels in the cells. After one hour of recovery in the dark, the cells from these three time points were divided in two equal aliquots. The first sample was directly re-sporulated as earlier described (Hersbach, GJM, Van der Beek, CP and Van Dijck, PWM. The Penicillins: properties, biosynthesis and fermentation. In: Vandamme EJ (ed) Biotechnology of Industrial Antibiotics (pp 45-140). Marcel Dekker, New York) and the other sample was incubated for an extended recovery period in YNB medium (0.67% w/v Yeast Nitrogen Base with amino acids (Difco), 2.0% w/v glucose) for 4 hours at 25 C before sporulation was induced.

Third mutagenized samples were obtained by germinating wild type spores overnight in YNB, followed by two washing steps in sterile tap water and resuspended in sterile tap water. Again UV irradiation was applied for 7.5, 15 and 30 minutes while the suspensions were slowly shaken. These samples were directly re-sporulated (as described above) after one hour of recovery in the dark.

To select the wanted mutants from these mutagenesized populations, the mutagenesized populations were inoculated in YEPD medium. After germination the development of cells was followed using standard light microscopy. When the average hyphae of a culture was nicely developing, cells were harvested and incubated with lysing enzymes to obtain protoplasts. Protoplasts were transformed with two DNA fragments carrying expression cassettes of functional selection markers, *ble* and *amdS*. The gene *ble* encodes a protein capable of conferring resistance to phleomycin (Kolar M, Punt PJ, van den Hondel CA, Schwab H. Transformation of *Penicillium chrysogenum* using dominant selection markers and expression of an *Escherichia coli* lacZ fusion gene. Gene. 1988;62(1):127-34). The gene *amdS* encodes a protein enabling cells to grow on acetamide as the sole nitrogen source (as described in EP635 574B). Techniques applied for the transfer of DNA to protoplasts of *P. chrysogenum* are well known in the art and are described in many references, including Finkelstein and Ball (eds.), Biotechnology of filamentous fungi, technology and products, Butterworth-Heinemann (1992); Bennett and Lasure (eds.) More Gene Manipulations in fungi, Academic Press (1991); Turner, In: Pühler (ed), Biotechnology, second completely revised edition, VHC (1992). The Ca-PEG mediated protoplast transformation is used as described in EP635574.

To select targeted integration of these expression cassettes to specific loci in the *Penicillium* genome short homologous stretches of DNA were added via PCR on both sides of the DNA fragments. Three types of construct were made: the first type contains homologous stretches of DNA of 30 bp, the second of 50 bp and the third of 100bp. Selection was performed transforming mutants obtained from the nine sporulated batches with two DNA constructs (*ble* and *amdS*) with 30, 50 or 100 bp homologous strectches defining 27 distinct batches.The *ble* gene was targeted to the *niaD* locus, thereby disrupting the nitrate reductase gene (Gouka RJ, van Hartingsveldt W, Bovenberg RA, van den Hondel CA, van Gorcom RF. Cloning of the nitrate-nitrite reductase gene cluster of *Penicillium chrysogenum* and use of the *niaD* gene as a homologous selection marker. J Biotechnol. 1991 Sep;20(2):189-99), enabling direct selection of transformants on plates containing chlorate, as cells become resistant to chlorate. The *amdS* gene was targeted to the *sutB* locus, thereby disrupting the sulphate permease gene (van de Kamp M, Pizzinini E, Vos A, van der Lende TR, Schuurs TA, Newbert RW, Turner G, Konings WN, Driessen AJ. Sulfate transport in *Penicillium chrysogenum:* cloning and characterization of the *sutA* and *sutB* genes. J. Bacteriol. 1999 Dec;181(23):7228-34), enabling direct selection of transformants on plates containing selenate. Transformants were first selected on chlorate and then tested for selenate. Furthermore, the presence of the selection markers was demonstrated by growth on plates containing acetamide as sole nitrogen source (EP635574B) and subsequently on plates containing phleomycine. As control wild type *P. chrysogenum* CBS 455.95 was also transformed with the same DNA fragments. Mutants with both selection markers present and resistant against both chlorate and selenate are strains with improved targeted integration.